(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 2 227 775 B1

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention of the grant of the patent:
17.12.2014 Bulletin 2014/51

(21) Application number: **08867791.9**

(22) Date of filing: **24.11.2008**

(51) Int Cl.:
G06K 9/20 *(2006.01)* ***G06K 7/10*** *(2006.01)*
G10L 13/00 *(2006.01)*

(86) International application number:
PCT/US2008/084548

(87) International publication number:
WO 2009/085484 (09.07.2009 Gazette 2009/28)

(54) AUTOMATIC DOMINANT ORIENTATION ESTIMATION IN TEXT IMAGES BASED ON STEERABLE FILTERS

AUTOMATISCHE SCHÄTZUNG DER DOMINANTEN AUSRICHTUNG IN TEXTABBILDUNGEN AUF GRUNDLAGE VON STEUERBAREN FILTERN

ESTIMATION BASÉE SUR DES FILTRES ORIENTABLES D'UNE ORIENTATION DOMINANTE AUTOMATIQUE DANS DES IMAGES DE TEXTE

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MT NL NO PL PT RO SE SI SK TR

(30) Priority: **20.12.2007 US 961648**

(43) Date of publication of application:
15.09.2010 Bulletin 2010/37

(73) Proprietor: **Intel Corporation**
Santa Clara, CA 95054 (US)

(72) Inventor: **NESTARES, Oscar**
Sunnyvale
California 94085 (US)

(74) Representative: **Beresford, Keith Denis Lewis et al**
Beresford & Co.
16 High Holborn
London
WC1V 6BX (GB)

(56) References cited:
JP-A- 2006 319 416
KR-A- 19980 016 572
KR-A- 19990 058 744
US-A- 6 151 423
US-A1- 2003 086 721
US-A1- 2005 212 925

- **FREEMAN W T ET AL: "THE DESIGN AND USE OF STEERABLE FILTERS", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE SERVICE CENTER, LOS ALAMITOS, CA, US, vol. 13, no. 9, 1 September 1991 (1991-09-01), pages 891-906, XP000231520, ISSN: 0162-8828, DOI: DOI: 10.1109/34.93808**
- **SIMONCELLI E P ET AL: "The steerable pyramid: a flexible architecture for multi-scale derivative computation", PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING. (ICIP). WASHINGTON, OCT. 23 - 26, 1995; [PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING. (ICIP)], LOS ALAMITOS, IEEE COMP. SOC. PRESS, US, vol. 3, 23 October 1995 (1995-10-23), pages 444-447, XP010197217, DOI: DOI:10.1109/ICIP.1995.537667 ISBN: 978-0-7803-3122-8**

EP 2 227 775 B1

Description

BACKGROUND

[0001] Image processing systems may be utilized by visually impaired persons to obtain images of text and to process the text images so that the corresponding text can be audibly read aloud to the user by an appropriate text-to-speech engine. In such image processing systems, optical character recognition engines may be utilized to convert text in an image format into text in a character based format for further processing by the text-to-speech engine. However, often the text may not be optimally aligned with respect to horizontal in the image to be processor, or the image itself may not have been captured in a sufficient alignment with respect to horizontal, thereby causing the text to be rotationally offset from horizontal. Typically, an optical character recognition engine may require the text in the image to be aligned within a certain range, typically +/- 10 degrees from horizontal, in order to be able to properly recognize the characters contained in the text.

[0002] United States Patent No. 6151423 A (Melen) concerns optical character recognition and adjustment for misalignment of text at a discrete angle, for example adjustment in the case of a document being located upside-down relative to a scanner. The correct orientation for documents scanned by an OCR system is determined from the confidence factors associated with multiple character images identified in the document. A disclosed orientation determination module for determining the correct page orientation includes a confidence factor values buffer, a comparison module, a reference value buffer, a sort module, and a decision module. The orientation determination module obtains arrays of confidence factor values that respectively correspond to first and second page orientations (e.g. true orientation and upside-down orientation, respectively). The confidence factor values buffer stores and indexes the confidence factor values, the sort module sorts the values, and the reference value buffer stores comparison information such as threshold levels. The comparison module compares the first and second arrays of confidence factor values, such as by assessing selected values in the arrays, averaging such selected values, and determining whether and how many values exceed threshold levels. The decision module determiners the correct, or proper, document orientation.

[0003] United States Patent Application Publication US 2005/0071167 A1 (Levin et al.) concerns optical character recognition of text, in a text to speech conversion system. The orientation of the scanned text in the object field of the scanning camera may be arbitrary and an adjustment for this is made in the course of image processing. The text orientation is determined in order to make this adjustment. An image analysis module is used to process the captured image by globally rotating the image using a skew adjustment module so that the boundaries of the textual region of the scanned image are aligned in a desired manner with respect to the obtained image. As an example, the boundaries of a rectangular page are identified and substantially aligned in an orthogonal orientation. A variety of techniques may be used to modify the skew. One proposed technique includes converting the image to greyscale, and then applying an edge detection filter to the image. Once the textual regions are identified, the skew is modified by rotating the image.

[0004] An example of skew adjustment, albeit in the field of texture recognition, is given in the following article by Greenspan et al., "Rotation Invariant Texture Recognition Using Steerable Pyramid", IEEE Proceedings of the 12th IAPR International Conference on Pattern Recognition, Volume II, Confrence B: Pattern Recognition and Neural Networks, pp. 162-167, 1994. This document discloses a rotation-invariant texture recognition system which achieves rotation-invariant classification on a large database of textures as well as extraction of the rotation of the input image relative to the prestored texture database. Rotation-invariance is achieved by extracting rotation-invariant features. Use is made of a steerable pyramid of oriented filters and a Discrete Fourier Transform (DFT) for interpolating the steerable filter outputs to extract location-invariant features.

[0005] Examples of steerable filters are disclosed in the following article by William T Freeman, et al., "The Design and Use of Steerable Filters", IEEE Transactions on Patent Analysis and Machine Intelligence 13 (1991) September, No. 9, New York, US. This article discloses an efficient architecture to synthesise filters of arbitrary orientations from linear combinations of basis filters, allowing one to adaptively "steer" a filter to any orientation, and to determine analytically the filter output as a function of orientation.

SUMMARY OF THE INVENTION

[0006] The present invention provides an image processing method of the kind, such as that disclosed by US 2005/0071167 A1, comprising: obtaining image data of an image containing text, the text being disposed at an arbitrary angle of rotation with respect to horizontal detecting the angle of rotation of the text with respect to horizontal; rotating the image data by an amount based at least in part on said detecting to provide a rotated image to align the text to horizontal; and performing optical character recognition of the text on the rotated image.

[0007] In accordance with the present invention, the image processing method disclosed herein is characterised in that the detecting the angle of rotation of the text with respect to horizontal is performed by: processing the image data with a bank of separable derivative filters; filtering the processed image data with oriented filters tuned to two or more different orientations; demodulating the filtered image data, by demodulator means, to obtain energy samples of the filtered and processed image data; and interpolating the energy samples, by interpolator means, from which interpolated energy samples an orientation of the text may be determined based at least in part on a maximum energy value of the interpolating.

[0008] The present invention also provides an image processing apparatus of the kind, such as that disclosed by US 2005/0071167 A1, comprising: an image data block capable of obtaining image data of an image containing text, the text being disposed at an arbitrary angle of rotation with respect to horizontal; an orientation corrector capable of detecting the angle of rotation of the text with respect to horizontal, and being capable of rotating the image data by an amount based at least in part on said detecting to provide a rotated image to align the text to horizontal; and an optical character recognition engine capable of performing optical character recognition of the text on the rotated image.

[0009] In accordance with the present invention, the image processing apparatus disclosed herein is characterised by the orientation corrector comprising: a bank of separable derivative filters to filter the image data; oriented filters, tuned to two or more different orientations, to receive the filtered image data from the bank of separable derivative filters; a demodulator to demodulate the oriented filtered image data; a magnitude calculator to obtain energy samples of the image data from the demodulator; and an interpolator to interpolate the energy samples from which an orientation of the text may be determined based at least in part on a maximum energy value of the interpolation.

DESCRIPTION OF THE DRAWfNG FIGURES

[0010] Claimed subject matter is particularly pointed out and distinctly claimed in the concluding portion of the specification. However, such subject matter may be understood by reference to the following detailed description when read with the accompanying drawings in which:

FIG. 1 is a block diagram of an image processing system in accordance with one or more embodiments;

FIG. 2 is a block diagram of an orientation corrector capable of detecting and correcting an orientation of an image received by an image processing system in accordance with one or more embodiments;

FIG. 3 is a plot of energy versus orientation utilized to determine an orientation of an image in an orientation corrector in accordance with one or more embodiments;

FIG. 4 is an example image that has been rotated by an orientation corrector of an image processing system in accordance with one or more embodiments; and

FIG. 5 is a block diagram of an information handling system capable of tangibly embodying an imaging system having an orientation corrector in accordance with one or more embodiments.

[0011] It will be appreciated that for simplicity and/or clarity of illustration, elements illustrated in the figures have not necessarily been drawn to scale. For example, the dimensions of some of the elements may be exaggerated relative to other elements for clarity. Further, if considered appropriate, reference numerals have been repeated among the figures to indicate corresponding and/or analogous elements.

DETAILED DESCRIPTION

[0012] In the following detailed description, numerous specific details are set forth to provide a thorough understanding of claimed subject matter. However, it will be understood by those skilled in the art that claimed subject matter may be practiced without these specific details. In other instances, well-known methods, procedures, components and/or circuits have not been described in detail.

[0013] In the following description and/or claims, the terms coupled and/or connected, along with their derivatives, may be used. In particular embodiments, connected may be used to indicate that two or more elements are in direct physical and/or electrical contact with each other. Coupled may mean that two or more elements are in direct physical and/or electrical contact. However, coupled may also mean that two or more elements may not be in direct contact with each other, but yet may still cooperate and/or interact with each other. For example, "coupled" may mean that two or more elements do not contact each other but are indirectly joined together via another element or intermediate elements. Finally, the terms "on," "overlying," and "over" may be used in the following description and claims. "On," "overlying," and "over" may be used to indicate that two or more elements are in direct physical contact with each other. However, "over" may also mean that two or more elements are not in direct contact with each other. For example, "over" may mean that one element is above another element but not contact each other and may have another element or elements in between the two elements. Furthermore, the term "and/or" may mean "and", it may mean "or", it may mean "exclusive-or", it may mean "one", it may mean "some, but not all", it may mean "neither", and/or it may mean "both", although the scope of claimed subject matter is not limited in this respect. In the following description and/or claims, the terms "comprise" and "include," along with their derivatives, may be used and are intended as synonyms for each other.

[0014] Referring now to FIG. 1, a block diagram of an image processing system in accordance with one or more embodiments will be discussed. In one or more embodiments, image processing system 100 may be utilized to capture or receive an image that may include one or more passages or regions of text or writing, convert the text image data into text data, and then convert the text into an audible speech format, for example so that a user of image processing system 100 may be able to audile hear the words, phrases, and/or sentences in the image. For example, image processing system 100 may be utilized to assist a user having visual impairment and thus difficulty in seeing and/or reading. Image processing system 100 may be utilized by such a visually impaired user to read aloud words, phrases, sentences or other text captured or received in an image format, so that the user may hear the words, phrases, sentences, and/or other text as speech. Such speech may be electronically synthesized for example in one or more embodiments. As shown in FIG. 1, image processing system 100 may receive an image input 110 that may be provided to orientation corrector 112 for detecting and orienting an image received by image input 110. In one or more embodiments, image input 110 may comprise circuitry capable of capturing an image and converting the image into a digital format. For example, image input 110 may comprise a camera based on a charge coupled device (CCD) array, a complementary metal-oxide semiconductor (CMOS) array, or the like type of imaging device. Alternatively, image input 110 may comprise a front end capable of receiving an image in a digital format from another device that captured the image, and/or optionally converting the image data into an image format suitable for processing by image processing system 100. For example, in such embodiments, image input 110 may comprise a card or disk reader capable of reading image data from a memory card or disk, a video input system capable of capturing and/or receiving an input signal from another device providing a digital output signal, and so on. However, these are merely example embodiments for image input 110, and the scope of the claimed subject matter is not limited in these respects.

[0015] Orientation corrector 112 received image data from image input 110 corresponding to an image and/or a video stream. Orientation corrector 112 is utilized to preprocess the image data prior to processing by optical character reader (OCR) engine 114. Typically, OCR engine 114 may perform optimally if the text is presented to optical character engine 114 in a horizontal format, or oriented as nearly horizontal as possible. However, images captured and/or received by image processing system 100 may not always be received in a horizontal alignment, for example if a camera used to capture the text image is not well aligned with the camera. Likewise, text sometimes may be intentionally written non-horizontally, for example the text may be written vertically from bottom to top such as might appear on a poster or sign. In one or more embodiments, OCR engine 114 may be capable of detecting the orientation of the text with respect to a horizontal line, and then correct the image of the text by rotating the text by an appropriate angle so that the text falls within a horizontal alignment, and/or a sufficiently horizontal alignment, to provide a more optimal image for processing by OCR engine 114. The output of OCR engine 114 may then be provided to a text-to-speech synthesizer 116 for converting the text processed by OCR engine 114 into audible speech that may be heard by the user of image processing engine. It should be noted that any one or more of image input 110, orientation corrector 112, OCR engine 114, and/or text-to-speech block 116 may be implemented in whole or in part in hardware and tangibly embodied by an appropriate circuit, and/or implemented in whole or in part and tangibly embodied by an appropriate software module or program, and the scope of the claimed subject matter is not limited in this respect. Further details of the operation of orientation corrector are shown in and described with respect to FIG. 2, below.

[0016] Referring now to FIG. 2, a block diagram of an orientation corrector capable of detecting and correcting an orientation of an image received by an image processing system in accordance with one or more embodiments will be discussed. As shown in FIG. 2, the blocks of orientation corrector may be implemented to the image data 210 received from image input 110 of FIG. 1. In one or more embodiments, the operations performed by the blocks of orientation corrector may be applied to a multi resolution pyramid of the input image data 210, in which each level may be obtained by first low pass filtering the image data 210 from a previous level with a separable cubic b-spline filter, and then subsampling a factor of 2 in both dimensions. The input image from each level, using intensity values, is first filtered at separable derivatives block 212 with a bank of separable derivative filters, which may comprise Gaussian derivative filters for example. In the particular embodiment of separable derivatives block 212 shown in FIG. 2, 6^{th} order derivatives may be utilized to provide a suitable compromise between angular resolution and computation, since higher orders may provide better resolution but at a higher coast. The bank of filters implemented by separable derivatives block 212 may comprise first passing the set of derivatives in the horizontal (X) dimension from the Gaussian kernel to the sixth order derivative of this kernel. Subsequently, the set of derivatives may be applied to the vertical (Y) dimension in reverse order so that the resulting images comprise the set of two-dimensional (2D) sixth order partial derivatives of the input image data 210. To obtain the set of seven directional derivatives it may be sufficient to apply a set of seven appropriate linear combinations via oriented filters block 214 over the output of the previous stage of separable derivates block 212. From the set of directional derivatives provided by oriented filters block 214, demodulation and low pass filter block 216 may be utilized to obtain energy measurements by first demodulating the filter response by the tuning frequency of the filter, and then low pass filtering to eliminate the replicas of the spectrum. The magnitude of the complex output of demodulation and low pass filter block 216 may then be computed. In the case of global orientation estimation, the local energy measurements may be collapsed into a single global number for each oriented filter in each resolution level. This may be accomplished by computing the median across all the pixels in each energy image via median block 220, and then taking the average of the resulting values across different resolutions for the same orientation (not shown in Figure 2). The output of the median block 220 produces a set of energy samples, which comprise 7 energy samples in the present example, which are plotted as data points (stars) in the graph 300 of FIG. 3 which represents energy versus orientation. From these data points an interpolated curve of the energy versus orientation may be obtained by interpolation (steering) block 222, interpolating appropriately the responses from the seven samples to arrive at a continuous curve. In one or more embodiments, theoretical interpolation formulas for steerable filters may be utilized, although other interpolation methods may likewise be utilized, and the scope of the claimed subject matter is not limited in this respect. Finally, from the interpolation curve the orientation that produces a maximum energy response may be selected via max block 224 to determine the orientation of the text in the image data 210. In one or more embodiments, the text in image 210 is disposed at an orientation angle corresponding to the angle at maximum energy from the interpolation curve, plus 90 degrees clockwise, which may be indicated by orientation output 226. Orientation corrector 112 may then rotate the image data 210 by an amount of angular rotation indicated by orientation output so that the text may be aligned, or nearly aligned, with respect to horizontal using an appropriate interpolation technique. In one or more embodiments, such an interpolation technique may comprise a bilinear interpolation, however the scope of the claimed subject matter is not limited in this respect.

[0017] Referring now to FIG. 3, a plot of energy versus orientation utilized to determine an orientation of an image in an orientation corrector in accordance with one or more embodiments will be discussed. Plot 300 provides energy in arbitrary units (a.u.) on a vertical axis 310 versus orientation of the directional derivative filter in radians on a horizontal axis. As discussed with respect to FIG. 2, above, interpolation (steering) block 222 may provide approximate measured values for energy versus orientation of the corresponding directional derivative filter applied to image data 210, indicated as plot 314. From the set of 7 measured sample values, a continuous interpolation curve 316 of energy versus orientation may be obtained by applying appropriate steerable filter interpolation function to the samples. The interpolation curve may then be applied to max block 222 to obtain orientation an orientation output 226 that is representative of the orientation of the image contained in image data 210. The orientation output 226 allows orientation corrector to rotationally align the image with a horizontal line, for example as shown in and described with respect to FIG. 4, below.

[0018] Referring now to FIG. 4, an example image that has been rotated by an orientation corrector of an image processing system in accordance with one or more embodiments will be discussed. FIG. 4 represents an example image 400 that has been rotated by orientation corrector 112 to be aligned, or nearly aligned, with the horizontal in order to optimize the performance of OCR engine 114 to recognize the characters of text contained within image 400. The orientation detection and correction function provided by orientation corrector 112 may provide image preprocessing for image processing system 100, for example to perform an automatic OCR of pictures taken under non-controlled conditions, such as by low vision or blind persons. As described herein, orientation corrector 112 may be capable automatically correcting for in-plane rotation to an amount suitable for OCR engine 114, for example aligned to horizontal within +/- 10 degrees. Furthermore, orientation corrector 112 is capable of applying such an orientation detection and correction function to a local portion of an image, so that each local portion may be individually rotated by an amount suitable for OCR processing. The ability of orientation corrector 112 to perform local orientation correction for portions of an image may be obtained via the utilization of steerable filters via oriented filters block 214 since the filter responses are local, meaning that at each pixel there is a set of samples of the local Fourier energy. Therefore, it is possible for orientation corrector 112 to perform a local analysis instead of a global one, thereby allowing for the automatic detection of different regions containing text rotated to different orientations with respect to one another. In one embodiment, the detection of different local orientations may be performed by orientation corrector 112 by first performing clustering of the pixels, using for example k-means to cluster the vectors formed with the energy samples, and then estimating the orientation within each cluster. As a result, if the input image data 210 contains regions of text rotated at different orientations such local analysis can be applied to one or more of the local regions, although the scope of the claimed subject matter is not limited in this respect.

[0019] In one or more embodiments, a particular OCR engine 114 may be selected for image processing system 100 independent of the preprocessing capabilities of a given OCR engine 114 in terms of orientation estimation and correction, so that the selection may be made at least in part predominantly on the OCR performance of OCR engine 114 itself rather than being based on preprocessing capabilities, although the scope of the claimed subject matter is not limited in this respect.

[0020] Referring now to FIG. 5, a block diagram of an information handling system capable of tangibly embodying an imaging system having an orientation corrector in accordance with one or more embodiments will be discussed. Information handling system 500 of FIG. 5 may tangibly embody one or more of any of the components of imaging system 100 as shown in and described with respect to FIG. 1. For example, information handling system 500 may represent the hardware of orientation corrector 112, with greater or fewer components depending on the hardware specifications of the particular device or application. Although information handling system 500 represents one example of several types of computing platforms, information handling system 500 may include more or fewer elements and/or different arrangements of elements than shown in FIG. 5, and the scope of the claimed subject matter is not limited in these respects.

[0021] Information handling system 500 may comprise one or more processors such as processor 510 and/or processor 512, which may comprise one or more processing cores. One or more of processor 510 and/or processor 512 may couple to one or more memories 516 and/or 518 via memory bridge 514, which may be disposed external to processors 510 and/or 512, or alternatively at least partially disposed within one or more of processors 510 and/or 512. Memory 516 and/or memory 518 may comprise various types of semiconductor based memory, for example volatile type memory and/or non-volatile type memory. Memory bridge 514 may couple to a graphics system 520 to drive a display device (not shown) coupled to information handling system 500.

[0022] Information handling system 500 may further comprise input/output (I/O) bridge 522 to couple to various types of I/O systems. I/O system 524 may comprise, for example, a universal serial bus (USB) type system, an IEEE 1394 type system, or the like, to couple one or more peripheral devices to information handling system 500. Bus system 526 may comprise one or more bus systems such as a peripheral component interconnect (PCI) express type bus or the like, to connect one or more peripheral devices to information handling system 500. A hard disk drive (HDD) controller system 528 may couple one or more hard disk drives or the like to information handling system, for example Serial ATA type drives or the like, or alternatively a semiconductor based drive comprising flash memory, phase change, and/or chalcogenide type memory or the like. Switch 530 may be utilized to couple one or more switched devices to I/O bridge 522, for example Gigabit Ethernet type devices or the like. Furthermore, as shown in FIG. 5, information handling system 500 may include a radio-frequency (RF) block 532 comprising RF circuits and devices for wireless communication with other wireless communication devices and/or via wireless networks, although the scope of the claimed subject matter is not limited in this respect.

[0023] Although the claimed subject matter has been described with a certain degree of particularity, it should be recognized that elements thereof may be altered by persons skilled in the art without departing from the scope of claimed subject matter. It is believed that the subject matter pertaining to automatic dominant orientation estimation in text images based on steerable filters and/or many of its attendant utilities will be understood by the forgoing description, and it will be apparent that various changes may be made in the form, construction and/or arrangement of the components thereof without departing from the scope of the claimed subject matter or without sacrificing all of its material advantages, the form herein before described being merely an explanatory embodiment thereof, and/or further without providing substantial change thereto. It is the intention of the claims to encompass and/or include such changes.

Claims

1. An image processing method comprising:

obtaining image data of an image containing text, the text being disposed at an arbitrary angle of rotation with respect to horizontal;
detecting the angle of rotation of the text with respect to horizontal;
rotating the image data by an amount based at least in part on said detecting to provide a rotated image to align the text to horizontal; and
performing optical character recognition of the text on the rotated image;
characterised in that:

said detecting the angle of rotation of the text with respect to horizontal is performed by:

processing the image data with a bank (212) of separable derivative filters,
filtering the processed image data with oriented filters (214) tuned to two or more different orientations,
demodulating the filtered image data, by demodulator means (216), to obtain energy samples of the filtered and processed image data, and
interpolating the energy samples, by interpolator means (222), from which interpolated energy samples an orientation of the text may be determined based at least in part on a maximum energy value of the interpolating.

2. A method as claimed in claim 1, said detecting being performed by using about seven oriented filters tuned to different orientations.

3. A method as claimed in either claim 1 or 2, further comprising performing said detecting, said rotating, and said performing optical character recognition on two or more local regions in the image independently of one or more other regions.

4. A method as claimed in any preceding claim, further comprising:

applying a text-to-speech synthesis on an output of the optical character recognition to audibly reproduce the text.

5. A method as claimed in any preceding claim, wherein said rotating comprises:

aligning the text to within +/- 10 degrees with respect to horizontal.

6. A method as claimed in any preceding claim, said interpolating comprising applying an interpolation method for steerable filters to the energy samples.

7. A method as claimed in any preceding claim, said rotating comprising applying a bilinear interpolation to the image data.

8. An image processing apparatus, comprising:

an image data block (110) capable of obtaining image data of an image containing text, the text being disposed at an arbitrary angle of rotation with respect to horizontal;
an orientation corrector (112) capable of detecting the angle of rotation of the text with respect to horizontal, and being capable of rotating the image data by an amount based at least in part on said detecting to provide a rotated image to align the text to horizontal; and
an optical character recognition engine (114) capable of performing optical character recognition of the text on the rotated image;
characterised by:

the orientation corrector (112) comprising:

a bank (212) of separable derivative filters to filter the image data;
oriented filters (214), tuned to two or more different orientations, to receive the filtered image data from the bank (212) of separable derivative filters;
a demodulator (216) to demodulate the oriented filtered image data;
a magnitude calculator (220) to obtain energy samples of the image data from the demodulator; and
an interpolator (222) to interpolate the energy samples from which an orientation of the text may be determined based at least in part on a maximum energy value of the interpolation.

9. An apparatus as claimed in claim 8, said orientation corrector (112) comprising about seven oriented filters tuned to different orientations.

10. An apparatus as claimed in either claim 8 or 9, said orientation corrector (112) being capable of performing the detecting and rotating on two or more local regions in the image independently of one or more other regions.

11. An apparatus as claimed in any preceding claim 8-10, further comprising:

a text-to-speech synthesizer (116) capable of converting an output of the optical character recognition into an audible reproduction of the text.

12. An apparatus as claimed in any preceding claim 8-11, said orientation corrector (112) being capable of aligning the text to within +/- 10 degrees with respect to horizontal.

13. An apparatus as claimed in any preceding claim 8-12, the interpolator (222) being capable of applying an interpolation method for steerable filters to the energy samples.

Patentansprüche

1. Bildverarbeitungsverfahren, umfassend:

Erhalten von Bilddaten eines Text enthaltenden Bildes,
wobei der Text in einem beliebigen Drehwinkel bezüglich der Horizontalen angeordnet ist;
Detektieren des Drehwinkels des Texts bezüglich der Horizontalen;
Drehen der Bilddaten um einen Betrag, der mindestens teilweise auf dem Detektieren basiert, um ein gedrehtes Bild bereitzustellen, um den Text horizontal auszurichten; und
Durchführung von optischer Zeichenerkennung des Texts in dem gedrehten Bild;
dadurch gekennzeichnet, dass
das Detektieren des Drehwinkels des Texts bezüglich der Horizontalen durchgeführt wird durch:

Verarbeiten der Bilddaten mit einer Reihe (212) separierbarer Gradientenfilter,
Filtern der verarbeiteten Bilddaten mit orientierten Filtern (214), die auf zwei oder mehr verschiedene Orientierungen eingestellt sind,
Demodulieren der gefilterten Bilddaten mit Demodulationsmitteln (216), um Energieproben der gefilterten und verarbeiteten Bilddaten zu erhalten,
und
Interpolieren der Energieproben mit Interpolationsmitteln (222), wobei eine Orientierung des Texts aus den interpolierten Energieproben bestimmt werden kann mindestens teilweise basierend auf einem maximalen Energiewert der Interpolation.

2. Verfahren nach Anspruch 1, wobei die Detektion unter Verwendung von etwa sieben auf unterschiedliche Orientierungen eingestellten orientierten Filtern durchgeführt wird.

3. Verfahren nach Anspruch 1 oder 2, ferner umfassend die Durchführung des Detektierens, des Drehens und des Durchführens optischer Zeichenerkennung an zwei oder mehr lokalen Gebieten in dem Bild unabhängig von einem oder mehreren anderen Gebieten.

4. Verfahren nach einem der vorhergehenden Ansprüche, ferner umfassend:

Anwenden einer Text-zu-Sprache-Synthese an einer Ausgabe der optischer Zeichenerkennung, um den Text hörbar zu reproduzieren.

5. Verfahren nach einem der vorhergehenden Ansprüche, wobei das Drehen Folgendes umfasst:

Ausrichten des Texts auf innerhalb von ±10 Grad bezüglich der Horizontalen.

6. Verfahren nach einem der vorhergehenden Ansprüche, wobei das Interpolieren Anwenden eines Interpolationsverfahrens für Kompass-Filter auf die Energieproben umfasst.

7. Verfahren nach einem der vorhergehenden Ansprüche, wobei das Drehen Anwenden einer bilinearen Interpolation auf die Bilddaten umfasst.

8. Bildverarbeitungsvorrichtung, umfassend:

einen Bilddatenblock (110), der imstande ist, Bilddaten eines Text enthaltenden Bildes zu erhalten, wobei der Text in einem beliebigen Drehwinkel bezüglich der Horizontalen angeordnet ist;
einen Orientierungskorrektor (112), der imstande ist,
den Drehwinkel des Texts bezüglich der Horizontalen zu detektieren, und der imstande ist, die Bilddaten um einen Betrag zu drehen, der mindestens teilweise auf dem Detektieren basiert, um ein gedrehtes Bild bereitzustellen, um den Text horizontal auszurichten;
und
eine optische Zeichenerkennungsvorrichtung (114), die imstande ist, optische Zeichenerkennung an dem Text in dem gedrehten Bild durchzuführen;
dadurch gekennzeichnet, dass
der Orientierungskorrektor (112) Folgendes umfasst:

eine Reihe (212) separierbarer Gradientenfilter zum Filtern der Bilddaten;
orientierte Filter (214), die auf zwei oder mehr verschiedene Orientierungen eingestellt sind, um die gefilterten Bilddaten von der Reihe (212) separierbarer Gradientenfilter zu empfangen;
einen Demodulator (216) zum Demodulieren der orientierungsgefilterten Bilddaten;
einen Größenberechner (220), um Energieproben der Bilddaten von dem Demodulator zu erhalten; und
einen Interpolator (222) zum Interpolieren der Energieproben, wobei eine Orientierung des Texts aus den interpolierten Energieproben bestimmt werden kann mindestens teilweise basierend auf einem maximalen Energiewert der Interpolation.

9. Vorrichtung nach Anspruch 8, wobei der Orientierungskorrektor (112) etwa sieben auf unterschiedliche Orientierungen eingestellte orientierte Filter umfasst.

10. Vorrichtung nach entweder Anspruch 8 oder Anspruch 9, wobei der Orientierungskorrektor (112) imstande ist, das Detektieren und Drehen an zwei oder mehr lokalen Gebieten in dem Bild unabhängig von einem oder mehreren anderen Gebieten durchzuführen.

11. Vorrichtung nach einem der vorhergehenden Ansprüche 8-10, ferner umfassend:

einen Text-zu-Sprache-Synthesizer (116), der imstande ist, eine Ausgabe der optischen Zeichenerkennung in eine hörbare Reproduktion des Textes umzuwandeln.

12. Vorrichtung nach einem der vorhergehenden Ansprüche 8-11, wobei der Orientierungskorrektor (112) imstande ist, den Text auf innerhalb von ±10 Grad bezüglich der Horizontalen auszurichten.

13. Vorrichtung nach einem der vorhergehenden Ansprüche 8-12, wobei der Interpolator (222) imstande ist, ein Interpolationsverfahren für Kompass-Filter auf die Energieproben anzuwenden.

Revendications

1. Procédé de traitement d’image, comprenant les étapes suivantes :

obtenir des données d’image d’une image contenant du texte, le texte étant disposé suivant un angle de rotation arbitraire par rapport à l’horizontale ;
détecter l’angle de rotation du texte par rapport

à l'horizontale ;
faire pivoter les données d'image d'une quantité basée, au moins en partie, sur ladite détection pour délivrer une image tournée pour aligner le texte à l'horizontale ; et
exécuter une reconnaissance optique de caractères du texte sur l'image tournée ;
caractérisé en ce que :

ladite détection d'angle de rotation du texte par rapport à l'horizontale est exécutée selon les étapes suivantes :

traiter les données d'image avec une banque (212) de filtres différentiels séparables,
filtrer les données d'image traitées avec des filtres orientés (214) ajustés selon deux orientations différentes, ou plus,
démoduler les données d'image filtrées, au moyen d'un démodulateur (216), pour obtenir des échantillons d'énergie des données d'image traitées et filtrées, et
interpoler les échantillons d'énergie, au moyen d'un interpolateur (222), à partir desquels échantillons d'énergie interpolés une orientation du texte peut être déterminée sur la base, au moins en partie, d'une valeur d'énergie maximum de l'interpolation.

2. Procédé tel que revendiqué dans la revendication 1, ladite détection étant exécutée en utilisant environ sept filtres orientés ajustés selon différentes orientations.

3. Procédé tel que revendiqué dans la revendication 1 ou la revendication 2, comprenant en outre d'exécuter ladite détection, ladite rotation, et ladite exécution de reconnaissance optique de caractères sur deux, ou plus, régions locales dans l'image indépendamment d'une ou plusieurs autres régions.

4. Procédé tel que revendiqué dans l'une quelconque des revendications précédentes, comprenant en outre l'étape suivante :

appliquer une synthèse vocale à partir du texte sur une sortie de la reconnaissance optique de caractères pour reproduire, de façon audible, le texte.

5. Procédé tel que revendiqué dans l'une quelconque des revendications précédentes, dans lequel ladite rotation comprend l'étape suivante :

aligner le texte à +/- 10 degrés par rapport à l'horizontale.

6. Procédé tel que revendiqué dans l'une quelconque des revendications précédentes, ladite interpolation comprenant d'appliquer un procédé d'interpolation pour filtres orientables aux échantillons d'énergie.

7. Procédé tel que revendiqué dans l'une quelconque des revendications précédentes, ladite rotation comprenant d'appliquer une interpolation bilinéaire aux données d'image.

8. Appareil de traitement d'image, comprenant :

un bloc de données d'image (110) capable d'obtenir des données d'image d'une image contenant du texte, le texte étant disposé selon un angle de rotation arbitraire par rapport à l'horizontale ;
un correcteur d'orientation (112) capable de détecter l'angle de rotation du texte par rapport à l'horizontale, et étant capable de faire tourner les données d'image d'une quantité basée, au moins en partie, sur ladite détection pour délivrer une image tournée pour aligner le texte à l'horizontale ; et
un moteur de reconnaissance optique de caractères (114) capable d'exécuter une reconnaissance optique de caractères du texte sur l'image tournée ;
caractérisé par :

le correcteur d'orientation (112) comprenant :

une banque (212) de filtres différentiels séparables pour filtrer les données d'image ;
des filtres orientés (214) ajustés selon deux, ou
plus, orientations différentes, pour recevoir les données d'image filtrées depuis la banque (212) de filtres différentiels séparables ;
un démodulateur (216), pour démoduler les données d'image filtrées orientées ;
un calculateur d'amplitude (220) pour obtenir des échantillons d'énergie des données d'image depuis le démodulateur ; et
un interpolateur (222), pour interpoler les échantillons d'énergie à partir desquels une orientation du texte peut être déterminée sur la base, au moins en partie, d'une valeur d'énergie maximum de l'interpolation.

9. Appareil tel que revendiqué dans la revendication 8, ledit correcteur d'orientation (112) comprenant environ sept filtres orientés ajustés selon différentes orientations.

10. Appareil tel que revendiqué dans la revendication 8 ou la revendication 9, ledit correcteur d'orientation (112) étant capable d'exécuter les détection et rotation sur deux régions locales, ou plus, dans l'image indépendamment d'une ou plusieurs autres régions.

11. Appareil tel que revendiqué dans l'une quelconque des revendications 8 à 10, comprenant en outre un synthétiseur vocal à partir du texte (116) capable de convertir une sortie de la reconnaissance optique de caractères en une reproduction audible du texte.

12. Appareil tel que revendiqué dans l'une quelconque des revendications 8 à 11, ledit correcteur d'orientation (112) étant capable d'aligner le texte à +/- 10 degrés par rapport à l'horizontale.

13. Appareil tel que revendiqué dans l'une quelconque des revendications 8 à 12, l'interpolateur (222) étant capable d'appliquer un procédé d'interpolation pour filtres orientables aux échantillons d'énergie.

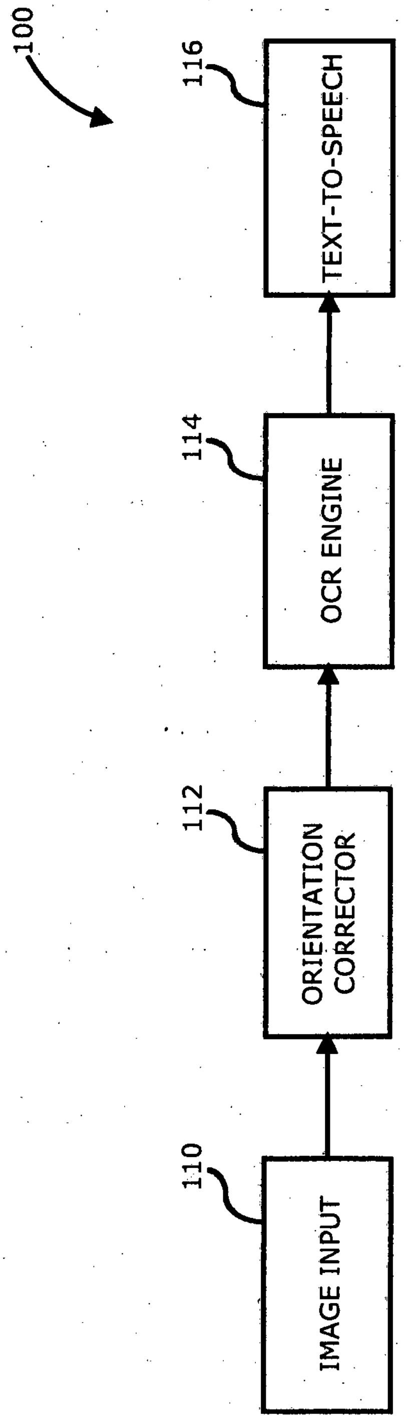
100
116
TEXT-TO-SPEECH
114
OCR ENGINE
112
ORIENTATION
CORRECTOR
110
IMAGE INPUT

FIG. 1

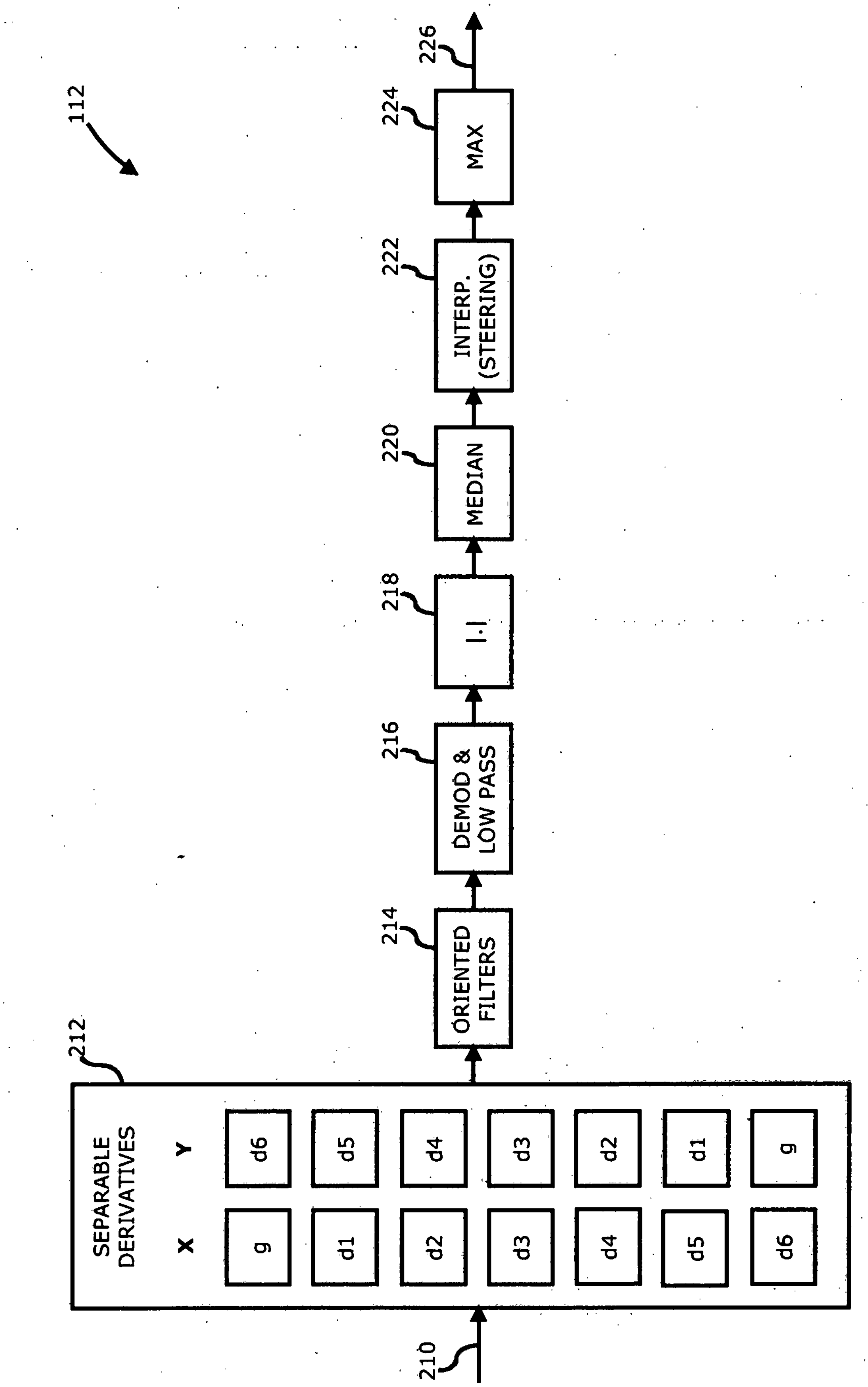
112
210
212
SEPARABLE DERIVATIVES
X
Y
g
d1
d2
d3
d4
d5
d6
214
ORIENTED FILTERS
216
DEMOD & LOW PASS
218
|.|
220
MEDIAN
222
INTERP. (STEERING)
224
MAX
226

FIG. 2

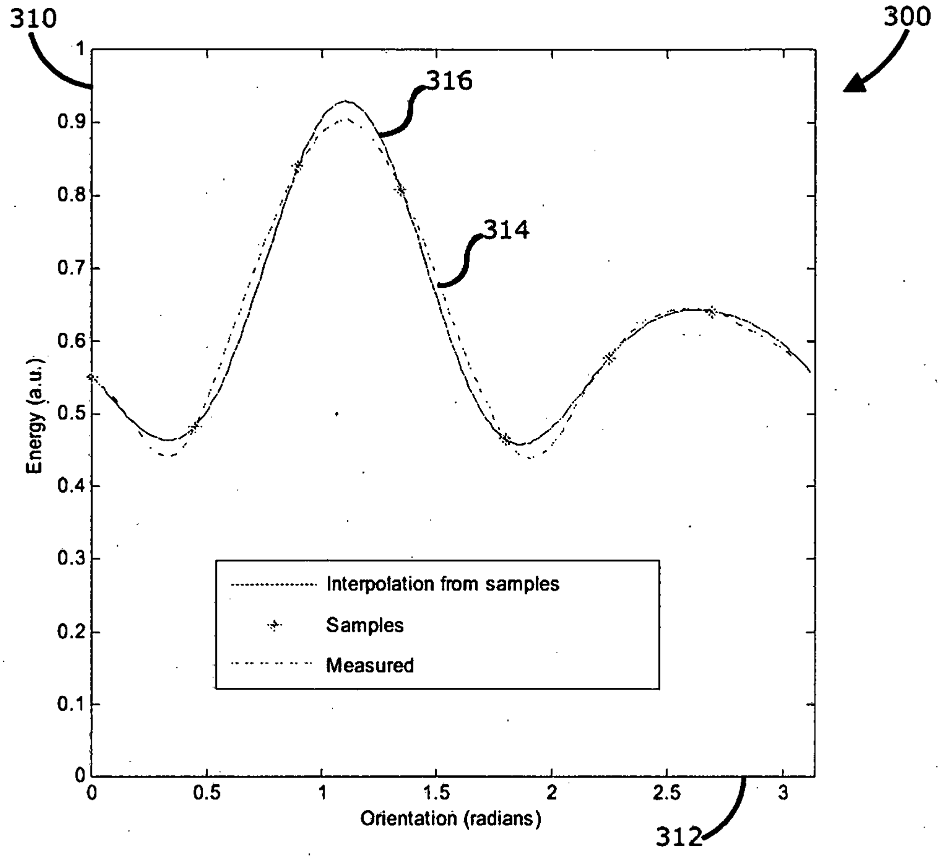
310
300
316
314
312
Energy (a.u.)
Orientation (radians)
Interpolation from samples
Samples
Measured
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
0.5
1
1.5
2
2.5
3

FIG. 3

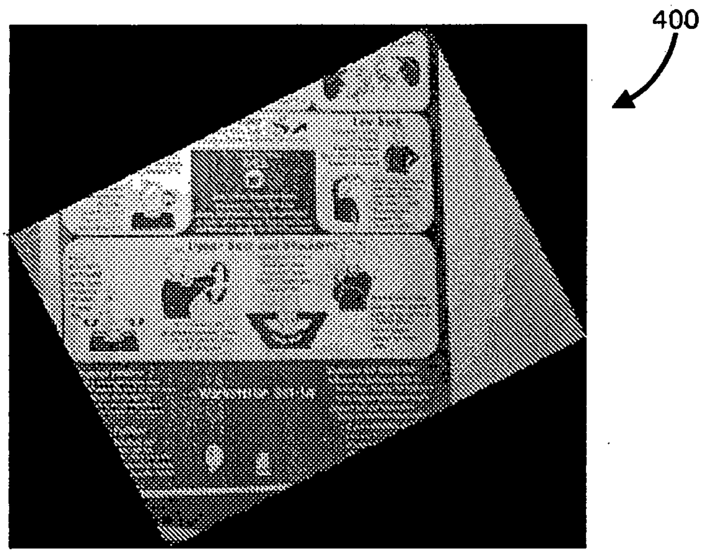
400

FIG. 4

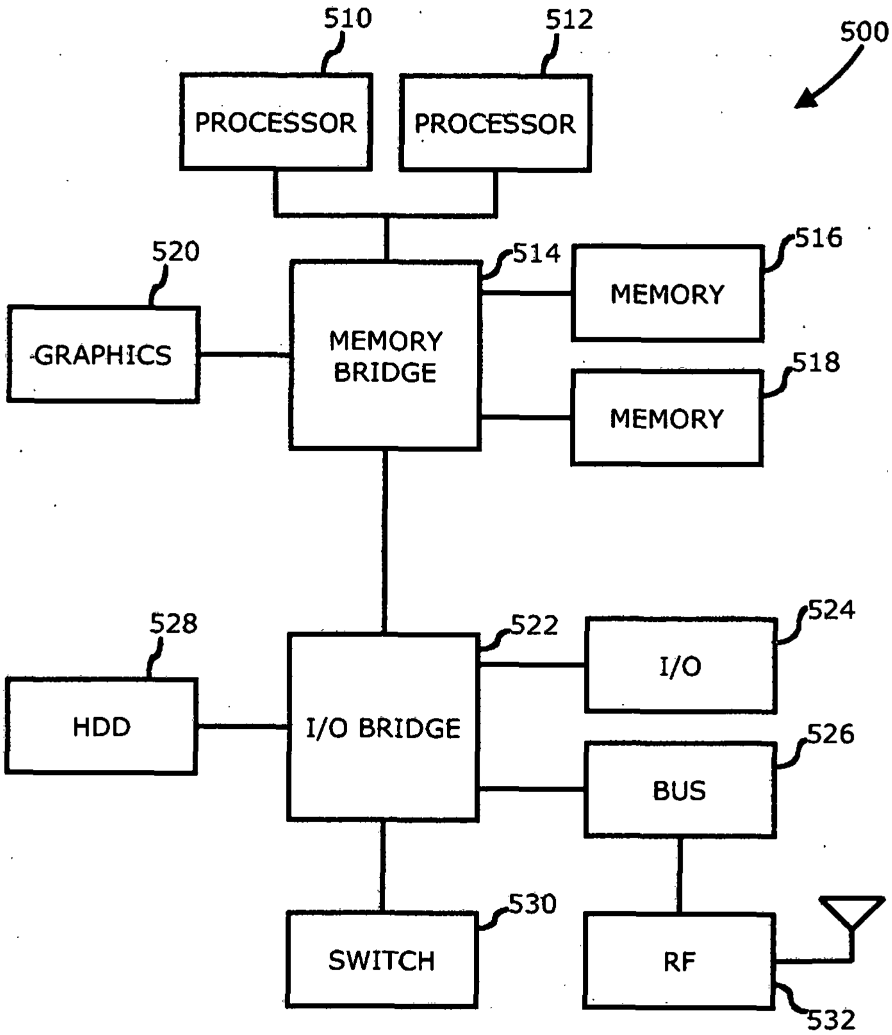
510
512
500
PROCESSOR
PROCESSOR
520
514
MEMORY
516
GRAPHICS
MEMORY
BRIDGE
518
MEMORY
528
522
I/O
524
HDD
I/O BRIDGE
526
BUS
530
SWITCH
RF
532

FIG. 5

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- US 6151423 A, Melen **[0002]**
- US 20050071167 A1, Levin **[0003] [0006] [0008]**

Non-patent literature cited in the description

- **GREENSPAN et al.** Rotation Invariant Texture Recognition Using Steerable Pyramid. *IEEE Proceedings of the 12th IAPR International Conference on Pattern Recognition, Volume II, Confrence B: Pattern Recognition and Neural Networks,* 1994, vol. II, 162-167 **[0004]**
- **WILLIAM T FREEMAN et al.** The Design and Use of Steerable Filters. *IEEE Transactions on Patent Analysis and Machine Intelligence,* September 1991, vol. 13 (9 **[0005]**